UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

MULTIPLE REGRESSION ANALYSIS WITH BOX-COX TRANSFORMATIONS AND NONSPHERICAL RESIDUAL ERRORS: A TRANSPORTATION APPLICATION

by

Marcel G. Dagenais
Marc J.I. Gaudry
Tran Cong Liem

This research was partly supported by the Research and Development Centre of Transport Canada and by the F.C.A.C. program of the *Ministère de l'Éducation du Québec*. Preliminary results were presented at the June 1979 Montreal Meeting of The Econometric Society.

Ce rapport a été publié grâce à une subvention du fonds F.C.A.C. pour l'aide et le soutien à la recherche.

Université de Montréal

Centre de recherche sur les transports - Publication # 166
Département de sciences économiques - Cahier # 8011

Mai 1980 - Révisé Avril 1982 et Décembre 1982

Dépôt légal - 2e trimestre 1980

ABSTRACT

In this paper, we develop an algorithm for solving regression models with Box-Cox transformations on both the dependent and the independent variables, while simultaneously taking into account corrections for serial correlation of several orders and for heteroscedasticity. The latter correction is of a general form which contains as special cases most specifications of heteroscedasticity encountered in practice. We apply the procedure to a transit demand function used in its linear form by the Montreal Transit Authority. Our results show that taking into account nonsphericalness of the residuals has a major impact on model parameters, notably on those which determine the functional form of the model, but a relatively minor impact on transit demand elasticities.

RÉSUMÉ

Nous formulons dans cet article une procédure d'estimation simultanée des paramètres dans un modèle de régression dont les variables dépendante et indépendantes sont assujetties à des transformations de Box-Cox. Notre procédure permet aussi de tenir compte de l'autocorrélation et de l'hétéroscédasticité des erreurs résiduelles. La correction de l'hétéroscédasticité est très générale et comprend comme cas particuliers la plupart des formulations courantes. Nous appliquons notre algorithme à un modèle de demande de transport en commun utilisé dans sa forme linéaire par la Commission de Transport de Montréal. Nous montrons l'influence réciproque majeure des paramètres qui décrivent la forme fonctionnelle et de ceux qui décrivent la distribution des erreurs. Les estimations des élasticités de la demande de transport en commun s'avèrent toutefois robustes par rapport au niveau de généralité du modèle.

CONTENTS

1. Introduction 2

2. The Model. 3

3. The Likelihood Function. 7

4. The Calculation of Elasticities. 12

5. The Algorithm. 16

6. Application to a Transit Demand Function 19

Reasonableness of the Stochastic Approach. 21

Behavior of the Likelihood Function. 22

Impact of Nonspherical Errors on the Functional Form . . . 22

Corrections for Autocorrelation and Heteroscedasticity . . 23

The Number of Functional Form Parameters 24

Impact of Nonspherical Errors on Elasticities of Transit Demand 24

Conclusion 25

References 29

Table I. Transit Demand by Montreal School Children (December 1956 - December 1971). 27

MAXIMUM LIKELIHOOD ESTIMATION OF A MULTIPLE REGRESSION MODEL WITH BOX-COX TRANSFORMATIONS AND NONSPHERICAL RESIDUAL ERRORS *

1. INTRODUCTION

The choice of the appropriate functional form is a crucial step in applied regression analysis. Theil (1971) has pointed out the importance of discriminating between linear and log-linear models for forecasting. A useful compromise has been suggested by Box and Cox (1964): applied both to the dependent and the independent variables, the Box-Cox transformation (BCT) defines a general class of models which includes the pure linear and log-linear models as special cases. The flexibility of the BCT has made it attractive for an increasing number of applications, for instance in monetary economics [Zarembka (1968), Spitzer (1976, 1977)], income analysis [Heckman and Polachek (1974), Welland (1978)], production theory [Appelbaum (1979), Berndt and Khaled (1979)] and transportation [Kau and Sirmans (1976), Gaudry (1980a), Wang et al. (1981)].

However, the application of the BCT, particularly to the dependent variable of regression models, does not merely affect the functional form of the model. It also affects at the same time the distributional properties of the residual errors. Box and Cox (1964) mention that the use of the BCT may render the density of the error term more normal-like; Zarembka (1974) points out that using the BCT may compensate for heteroscedasticity. Furthermore, attention has recently been paid to the impact of first order [Savin and White (1978)] and multiple order [Gaudry and Wills (1978)] autocorrelation on the BCT parameter estimates.

---

* This research was partly supported by the Research and Development Centre of Transport Canada and by the F.C.A.C. program of the ministère de l'Education du Québec. Lina Di Pietro and Lisa Di Pietro made contributions as research assistants. Preliminary results were presented at the June 1979 Montreal Meeting of the Econometric Society.

The fact that the application of the BCT *simultaneously* affects the functional form of the regression model and the form of the density of the residual errors has disadvantages. For example, if one had a curvilinear regression model with homoscedastic disturbances, one might use the BCT to obtain a better fit but this action might simultaneously introduce undesired heteroscedasticity in the residual errors. Alternatively, one could have a linear model with heteroscedastic disturbances where the use of the BCT would reduce the heteroscedasticity but impair the linear fit.

In order to avoid such a dilemma, one should consider separate instruments for the different targets: the BCT to improve the functional fit and a different scheme to correct for heteroscedasticity. This strategy, suggested in Gaudry and Dagenais (1978, 1979), will be substantiated here and tested with an application to transit demand.

2. THE MODEL

In line with the above considerations, the regression model proposed in this paper will have the following form:

$$y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t \quad , \quad (t = 1,\ldots,N) \qquad (1)$$

$$u_t = [f(Z_{1t},\ldots,Z_{Mt})]^{1/2} v_t \quad , \qquad (2)$$

where

y_t is the dependent variable,

X_{kt} for k = 1,...,K, and Z_{mt} for m = 1,...,M, are independent variables[1],

$y_t^{(\lambda_y)}$ and $X_{kt}^{(\lambda_{xk})}$ denote variables affected by BCT, for instance

$$X_{kt}^{(\lambda_{xk})} = \frac{X_{kt}^{\lambda_{xk}} - 1}{\lambda_{xk}} \quad . \tag{3}$$

Furthermore, we have: $-\infty < v_t < +\infty$, $E(v_t) = 0$ and $E(v_t^2) = \psi^2$. The β's, λ's and ψ^2 are parameters. Clearly, BCT are applied only to strictly positive variables; other variables are left untransformed.

Since BCT on y_t are defined only for $y_t > 0$, *it is evident that equation (1) may not strictly hold for every value of* v_t. This point will be further discussed in the next Section.

A useful form that can be given to the general heteroscedasticity function shown in (2) is the following:

$$f(Z_{1t},\ldots,Z_{Mt}) = \{\lambda_u [\delta_o + \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}] + 1\}^{1/\lambda_u} \quad . \tag{4}$$

In this expression, direct BCT are applied to the Z variables which explain error variance and an inverse power transformation is applied to the expression in the square parentheses. The constant δ_o is necessary to preserve the invariance of the expression to units of measurement of the Z's [Schlesselman (1971)]. Equation (4) contains as special cases most schemes of traditional interest.

---

[1]. Note that the X and Z sets may have common elements.

The first family of special cases is based on setting $\lambda_u = 0$. This yields

$$E(u_t^2) = \omega_{tt} = \psi^2 \exp\{\delta_o + \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\}$$
$$= \phi^2 \exp\{\sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\} \qquad (5)$$

where

$$\phi^2 = \psi^2 \exp\{\delta_o\}.$$

Furthermore, if each λ_{zm} is also set equal to zero, the last expression of equation (5) reduces to the multivariate multiplicative model found in Dagum and Dagum (1974):

$$\omega_{tt} = \phi^2 \prod_{m=1}^{M} Z_{mt}^{\delta_m} \quad . \qquad (6)$$

Equation (6) finally reduces to the classical multivariate form:

$$\omega_{tt} = \phi^2 \prod_{m=1}^{M} Z_{mt}^2 \qquad (7)$$

when each δ_m is set equal to 2. Park's (1966) specification is obtained from (5) by setting all δ_m, but one, equal to zero and by setting the λ_{zm} of the retained variable equal to zero. Imposing the further constraint that the retained δ_m equals 2 yields the classical univariate form $\phi^2 Z_{mt}^2$. Egy and Lahiri (1979) have recently used Park's specification with an equation of form (1).

Another family of special cases is obtained by setting $\lambda_u = 1$. This results in:

$$\omega_{tt} = \tilde{\delta}_o + \sum_{m=1}^{M} \tilde{\delta}_m Z_{mt}^{(\lambda_{zm})} \qquad (8)$$

where $\tilde{\delta}_o = \psi^2 (\delta_o + 1)$, $\tilde{\delta}_m = \psi^2 \delta_m$. This family includes the linear form (each $\lambda_{zm} = 1$), such as that found in Glejser (1969), and many nonlinear forms (e.g. square root, if each $\lambda_{zm} = 0.5$ and second degree, if each $\lambda_{zm} = 2$) as special cases. Buse (1979) recently formulated a different definition of generalized heteroscedasticity which also includes the linear form as a special case.

The subclass defined by (8) may be less useful than the subclass defined by (5), because the former can produce negative values for the ω_{tt}'s. We shall therefore use the latter in the rest of this paper.

Since many of the data sets to which the model consisting of equations (1) to (5) will be applied in practice are time series, it is useful to assume that the v_t's follow a stationary r'th order autocorrelation process of the form:

$$v_t = \sum_{\ell=1}^{r} \rho_\ell v_{t-\ell} + w_t \qquad (9)$$

where w_t has zero mean, constant variance σ^2 and is independent among observations.

3. THE LIKELIHOOD FUNCTION

Assuming that the w_t's are normally distributed and neglecting the first r observations[2], the log-likelihood of the remaining N-r observations on the y_t's may be expressed as follows:

$$\mathcal{L} = \sum_{t=1+r}^{N} \mathcal{L}_t = -\frac{N-r}{2} \ln (2\pi\sigma^2) - \frac{1}{2} \sum_{t=1+r}^{N} \ln f(Z_t)$$

$$- \frac{1}{2\sigma^2} \sum_{t=1+r}^{N} \left[\frac{u_t}{\sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \frac{u_{t-\ell}}{\sqrt{f(Z_{t-\ell})}} \right]^2 + (\lambda_y - 1) \sum_{t=1+r}^{N} \ln y_t \qquad (10)$$

where

$$f(Z_t) = \exp \left\{ \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})} \right\} \qquad . \qquad (11)$$

Note that the last term of the RHS of (10) is the Jacobian of the transformation from the density of the $y_t^{(\lambda_y)}$'s to that of the y_t's and that the expression in the square brackets corresponds to w_t.

Since the BCT on y_t is defined only for $y_t > 0$, so that equation (1) can hold only for values of w_t which are consistent with this constraint, it might be thought that equation (10) is not strictly valid. We maintain, however, that *it is indeed the appropriate likelihood function*. Our argument is based on the fact that in very many econometric applications, it is reasonable to consider y_t as *limited* both *downwards* and *upwards*, even if the sample contains no limit observations.

---

2. Note that, when the number of observations is small, it might be advantageous to recuperate the first r observations, as pointed out by Beach and MacKinnon (1978). Recuperating the first r observations does not raise any theoretical problem, but the expressions of the log-likelihood function and of its first derivatives then become more messy: we have therefore not considered this alternative.

Several statistical interpretations of the BCT regression model are possible. For example, Bickel and Doksum (1981) assume $-\infty < y_t < \infty$, $\lambda > 0$ and use an extended definition of the power family, namely

$$y^{(\lambda)} = [|y|^{\lambda} \operatorname{sgn}(y) - 1]/\lambda \quad .$$

Bickel and Doksum note that "if in fact, the y's [in the sample] are positive, the likelihood function is as in Box and Cox". Adopting the Bickel-Doksum approach would therefore establish that equation (10) is indeed the appropriate likelihood function. Although Carroll and Ruppert (1981) are less explicit about their own assumptions, their use of the Bickel and Doksum asymptotic results suggests that they adopt essentially the same approach. However, for many econometric applications such as demand analysis, the assumption that y can take negative values is unrealistic; and for this reason, the Bickell-Doksum approach does not seem appropriate for our purposes.

Others, such as Spitzer (1976), Olsen (1978), Poirier (1978), Amemiya and Powell (1980), postulate that y_t cannot be negative and that the Box-Cox maximand is not strictly a likelihood function because it takes no account of the fact that y_t is truncated. This viewpoint is true if one (implicitly) assumes that the distribution function of $y_t^{(\lambda)}$ is everywhere continuous for all possible values of the explanatory variables, and consequently contains no mass point. We shall not adopt this approach because we believe that for some conceivable values of the explanatory variables, demand could very well take the value zero (or some value arbitrarily close to zero)

with non negligeable probability. In such cases, it is perfectly possible, for instance, for the distribution of $y_t^{(\lambda)}$ to have a mass point at $-1/\lambda$ when $\lambda > 0$, for particular values of the explanatory variables.

In a large number of econometric applications, such as demand analysis, it is indeed very reasonable to assume that the dependent variable has both a lower and an upper limit, even if these points are almost never actually observed. In the application presented in Section 6, the dependent variable is the quantity of bus and subway trips per month by school children in Montreal. If the transit fare were increased from its present value of 25 cents to 40 $ or 60 $ — which is considerably higher than taxi fares within the urban area —, it is reasonable to assert that the demand for bus travel by school children in Montreal would be reduced to zero. (In practice, it is convenient, as we shall see later, to use as lower limit ε, an arbitrarily small positive number, rather than zero because the BCT can only be effected on strictly positive variables since it includes the log transformation: $\lambda = 0$). On the other hand, if the transit fare were set at zero or even at a negative value, — which would imply that the Montreal Transit Authority pay children to use transit — the demand would surely increase considerably but could not increase strictly to infinity. Given the number of available transit vehicles and the number of school children in Montreal, the monthly travel demand by school children could surely never be larger than some theoretical finite upper limit, say L.

In fact, one could argue that most standard regression analyses made in economics and in many other fields pertain to doubly limited dependent variables. This fact is usually ignored because the samples contain no limit observations and the analyses and predictions usually concern values of y_t or $E(y_t)$ which are far from both the upper and the lower limits.

The proper statistical model to use is therefore the two-limit Tobit model [Rosett and Nelson (1975)].

If the true model is the Rosett-Nelson model and if the actual sample contains no limit observations, the *exact likelihood function* reduces to that suggested by Box and Cox for regression models with homoscedastic and serially uncorrelated error terms and *to our equation (10)* for models involving heteroscedasticity and serial correlation. This can be readily inferred by considering, for example, equation (10) of the Rosett-Nelson paper and setting the number of limit observations in the sample equal to zero.

Under the assumption that the true model of y_t is that of a two-limit Tobit model, and given that the actual sample available contains no upper or lower limit observations, *the parameter estimators that maximize equation (10) are the true maximum likelihood estimators*[3]. What then, are the asymptotic properties of these estimators ? Different possible cases must

---

[3]. Strictly speaking, they are conditional maximum likelihood estimators, conditional on the values of the first r observations. Similarly, when we speak of the probability that y_t be equal to its upper or lower limit, or when we discuss the calculation of $E(y_t)$, these are all *conditional* on the past values of y.

be considered here. In line with the remarks made by Draper and Cox (1969) about the original BCT model, if it can be safely assumed that the parameters of the model, the initial conditions and the empirical distribution of the exogenous variables are such that the probability that $y_t^{(\lambda)}$ be equal to its higher or lower limit is always negligible, it is conjectured that "the general large-sample theorems about the sampling distributions of maximum likelihood estimates", as Box and Cox (1969, p. 219) put it, apply to $\mathcal{L}_{max}$ of equation (10)[4]. Note that, as also suggested by Draper and Cox and reiterated by Olsen (1978), "the reasonableness of the [above] assumptions ... can be tested by an analysis of the residuals". For that purpose, the probability that $y_t^{(\lambda)}$ be equal to its lower limit or to its upper limit should be estimated for each observation in the sample, as illustrated in our application (see Section 6).

---

4. This point has also been recognized by other authors in the literature, for the original Box-Cox model. For example, Zellner (1971, p. 163, fn. 2) notes that the range of $y_t^{(\lambda)}$ is not $-\infty$ to $+\infty$ "but just a subinterval of this range". He then points out that if the probability that $y_t^{(\lambda)}$ "will fall in the excluded interval is small, the normality assumption will not be vitally affected". In the same line of thought, it is readily seen in Amemiya and Powell (1980, p. 6, equation 2.10) that, if it is assumed that $y_t^{(\lambda)}$ has a truncated normal distribution and if the surface under the normal curve for the admissible region is close to one, the asymptotic bias of the "Box-Cox MLE" of λ vanishes; in their notation, if

$$\Phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right) \to 1 \text{ for every } t \text{ , then } -\frac{1}{\sigma_o\lambda_o^2}\sum \frac{\phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)}{\Phi\left(\frac{1}{\lambda_o\sigma_o} + \frac{\mu_{to}}{\sigma_o}\right)} \text{ will} \to 0 \text{ .}$$

Clearly, a rigorous analytical derivation of the asymptotic properties of the proposed estimators of the model discussed in the present paper would require further investigation. An expensive alternative approach, which could generate information on the finite sample properties as well as on the asymptotic properties of the estimators, would be to resort to Monte Carlo experiments.

Even if it cannot be safely assumed that the probability that $y_t^{(\lambda)}$ reaches its lower or upper limit is always negligible, this probability may be sufficently small that the probability of observing samples without limit observations remains very high. Then, in such cases, our maximum likelihood estimators remain presumably consistent and asymptotically normally distributed, given that Amemiya has shown that the maximum likelihood estimates of the Tobit model possesses such asymptotic properties.[5]

Finally, if the probability that $y_t^{(\lambda)}$ be equal to its upper or lower limit is so high — given the values of the explanatory variables observed in the actual sample — as to make it unlikely that any sample of the size considered contains no limit observation, then one might clearly suspect that the suggested model is unreasonable. Note that, as pointed out above, the problem we are discussing at present is not specific to the BCT model, since it could occur with any linear regression model with a non negative dependent variable. Although this fact is most often neglected, it would be just as justified and imperative to examine the probability that the dependent variable be zero in every linear regression model with non negative dependent variable as it is for the BCT model.

4. THE CALCULATION OF ELASTICITIES

A related problem is that of obtaining elasticities of the dependent variable y_t, or more precisely of $E(y_t)$, with respect to the exogenous variables X_{it}. The evaluation of these elasticities defined, following

[5]. In Amemiya's (1973) paper, the proofs that are developed seem perfectly compatible with the fact that a given *finite* observable sample may contain zero limit observation. The difficulty of extending Amemiya's proof to the model discussed in the present paper comes mainly from the assumption of serial correlation. The introduction of serial correlation in limited dependent variable models complicates greatly the general expression of the likelihood function, since every run of m limit observations leads to the introduction of an m-dimensional integral of the multivariate normal density function [Dagenais (1982)]. However, Robinson's approach (1982) could possibly be used in this case.

Kmenta (1971), as $\eta\text{-}E = \frac{\partial E(y_t)}{\partial X_{it}} \cdot \frac{X_{it}}{E(y)}$, requires the computation of $E(y)$. Generalizing and extending Tobin's (1958) formula to this end, the expected value of the doubly limited dependent variable $\varepsilon \leq y_t \leq L$ is then given by:

$$E(y_t) = \varepsilon \int_{-\infty}^{\tilde{w}_t} f(w;\sigma^2)\, dw + \int_{\tilde{w}_t}^{w_t^*} y(w)_t\, f(w;\sigma^2)\, dw + L \int_{w_t^*}^{\infty} f(w;\sigma^2)\, dw \qquad (12)$$

where

$$\tilde{w}_t = \frac{\varepsilon^{\lambda_y} - 1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell\, y^*_{t-\ell} - \sum_{k=1}^{K} \beta_k\, X^{**}_{kt} \quad ,$$

$$w_t^* = \frac{L^{\lambda_y} - 1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell\, y^*_{t-\ell} - \sum_{k=1}^{K} \beta_k\, X^{**}_{kt} \quad ,$$

$$Z_t = (Z_{1t}, \ldots, Z_{Mt}) \quad ,$$

$$y^*_{t-\ell} = y^{(\lambda_y)}_{t-\ell} / \sqrt{f(Z_{t-\ell})} \quad ,$$

$$X^{**}_{kt} = X^{(\lambda_{xk})}_{kt} / \sqrt{f(Z_t)} - \sum_{\ell=1}^{r} \rho_\ell\, X^{(\lambda_{xk})}_{k,t-\ell} / \sqrt{f(Z_{t-\ell})} \quad ,$$

$$f(w;\sigma^2) = (2\pi)^{-1/2}\, \sigma^{-1} \exp\{-w^2/2\sigma^2\} \quad ,$$

and

$$y(w_t) = [1 + \lambda_y \sqrt{f(Z_t)}\, (\sum_{\ell=1} \rho_\ell\, y^*_{t-\ell} + \Sigma\, \beta_k\, X^{**}_{kt} + w_t)]^{1/\lambda_y} \quad .$$

Let us now successively examine, for $\lambda_y > 0$ and for $\lambda_y < 0$, how the three terms on the RHS of (12) behave as ε gets closer to zero and as L gets larger. These terms will be referred to, below, as expressions [1], [2] and [3] respectively from left to right.

When $\lambda_y > 0$, expression [1] $\to 0$ as $\varepsilon \to 0$. Similarly, expression [3] $\to 0$, as $L \to \infty$ since, for sufficiently large L, we necessarily have

$$L \int_{w_t^*}^{\infty} f(w;\sigma^2)\, dw \leq L \int_{w_t^*}^{\infty} w\, f(w;\sigma^2)\, dw \tag{13}$$

and

$$L \int_{w_t^*}^{\infty} w\, f(w;\sigma^2)\, dw \to 0 \text{ as } L \to \infty \quad .$$

Therefore, when both $\varepsilon \to 0$ and $L \to \infty$, we have:

$$E(y_t) = \int_{w_t^{**}}^{\infty} y(w)_t\, f(w;\sigma^2)\, dw \tag{14}$$

where

$$w_t^{**} = -\frac{1}{\lambda_y \sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell\, y_{t-\ell}^* - \sum_{k=1}^{K} \beta_k\, X_{kt}^{**} \quad .$$

The value of $E(y_t)$ can then be computed by numerical methods.

When $\lambda_y = 0$, expressions [1] and [3] both tend toward zero as $\varepsilon \to 0$ and $L \to \infty$. In addition term [2] can be rearranged to yield:

$$E(y_t) = \int_{-\infty}^{\infty} \bar{y}(w)_t\, f(w;\sigma^2)\, dw \tag{15}$$

where

$$\bar{y}(w)_t = \exp\{\sqrt{f(Z_t)}\, (\sum_{\ell=1}^{r} \rho_\ell\, y_{t-\ell}^* + \sum_{k=1}^{K} \beta_k\, X_{kt}^{**} + w_t)\} \quad .$$

Note therefore that, when $\lambda \geq 0$, the fact that y_t may have an upper limit is of no consequence in practice for the computation of $E(y_t)$, provided one assumes that this upper bound is sufficiently high. Indeed, for a sufficiently large value of L, the probability associated with the mass point at L tends toward zero for any finite values of X^{**}_{kt} $(k = 1,\ldots,K)$, $y^{*}_{t-\ell}$ $(\ell = 1,\ldots,r)$ and the elements of Z_t; furthermore, the value of w^{*}_{t} in [2] can be replaced by ∞ without altering the value of the integral. Therefore, when $\lambda_y \geq 0$, the upper bound L does not even have to be specified explicitly, provided it is assumed to be sufficiently large[6].

The case is different however for $\lambda_y < 0$. This case cannot be treated as the two previous cases since, if we assume that L get arbitrarily large and $\to \infty$, $E(y_t)$ does not exist. In this case, the value of L must be specified explicitly if one wants to compute $E(y_t)$. Therefore, when $\lambda < 0$, as $\varepsilon \to 0$, term [1] of (12) $\to 0$ and $E(y_t)$ becomes:

$$E(y_t) = \int_{-\infty}^{w^{*}_{t}} y(w)_t \, f(w;\sigma^2) \, dw + L \int_{w^{*}_{t}}^{\infty} f(w;\sigma^2) \, dw \quad . \qquad (16)$$

---

[6]. In certain cases, one may not be ready to make such an assumption and might prefer to specify the value of L explicitly. In such case, the expression for $E(y_t)$ would become:

$$\text{for } \lambda_y > 0 \; : \; E(y_t) = \int_{w^{**}_{t}}^{w^{*}_{t}} y(w)_t \, f(w;\sigma^2) \, dw + L \int_{w^{*}_{t}}^{\infty} f(w;\sigma^2) \, dw$$

and

$$\text{for } \lambda_y = 0 \; : \; E(y_t) = \int_{-\infty}^{\bar{w}_t} \bar{y}(w)_t \, f(w;\sigma^2) \, dw + L \int_{\bar{w}_t}^{\infty} f(w;\sigma^2) \, dw$$

where

$$\bar{w}_t = \frac{\ln L}{\sqrt{f(Z_t)}} - \sum_{\ell=1}^{r} \rho_\ell \, y^{*}_{t-\ell} - \sum_{\ell=1}^{K} \beta_k \, X^{**}_{kt} \quad .$$

Note that, for values of Z_t, X^{**}_t and y^*_t such that $E(y_t)$ is much smaller than L, the probability that y_t attain its upper limit L will be negligible and the value of $E(y_t)$ will be quite insensitive to the precise value retained for L. In samples where no upper limit values of y_t have been observed, one may infer that all the Z_t, X^{**}_t and y^*_t values contained in the sample are effectively such that $E(y_t)$ is much smaller than L for every t. Hence, it should be possible to compute $E(y_t)$ for the observations contained in the sample to the desired level of accuracy without requiring that the value of L be known with great precision: it should be sufficient to specify L only very approximately [7].

In Table 1, where results of our application are presented, both the "pseudo-(E)-R^2" and η-E, the elasticity of E(y), involve the computation of expected values from (14), (15) or (16). Elasticities are estimated at the average values of the independent variables and lagged y's.

5. THE ALGORITHM

In order to maximize $\mathcal{L}$ in equation (10) with respect to the vector of unknown parameters π, where $\pi = (\lambda_y, \lambda_{x1},\ldots,\lambda_{xK}, \lambda_{z1},\ldots,\lambda_{zM}, \delta_1,\ldots,\delta_M, \rho_1,\ldots,\rho_r, \beta_1,\ldots,\beta_K, \sigma^2)'$, the function is first concentrated over σ^2 and

---

[7]. Because of these considerations, in our program [Liem (1979)], the user is asked to specify the value of L. Otherwise, if the user *does not* specify L and the maximum likelihood estimate of λ_y turns out to be *negative*, the program generates, by default, its own approximation of L and uses it to compute $E(y_t)$ from equation (16). When the L value is thus generated by the program, it is automatically printed in the output, so that the user may be made aware of the approximation actually used. This procedure is further discussed in Liem (1979).

the β's. Then the concentrated likelihood function $\bar{\mathcal{L}}(\bar{\pi})$, where $\bar{\pi} = (\lambda_y, \lambda_{x1},\ldots,\lambda_{xK}, \lambda_{z1},\ldots,\lambda_{zM}, \delta_1,\ldots,\delta_M, \rho_1,\ldots,\rho_r)'$ is maximized with respect to the elements of $\bar{\pi}$, using the Davidon-Fletcher-Powell (DFP) algorithm [see Fletcher and Powell (1963)].

To obtain $\bar{\mathcal{L}}$, concentration over the β's is obtained by writing $\frac{\partial \mathcal{L}}{\partial \beta} = 0$, where $\beta = (\beta_1,\ldots,\beta_K)'$, and solving for β to get:

$$\bar{\beta} = (X^{**\prime}X^{**})^{-1} X^{**\prime} y^{**} \tag{17}$$

where

$$X^{**} = (X^{**\prime}_{r+1},\ldots,X^{**\prime}_{N}) \quad ,$$

$$X^{**}_{t} = (X^{**}_{1t},\ldots,X^{**}_{Kt}) \quad , \qquad (t = r+1,\ldots,N)$$

$$y^{**} = (y^{**}_{r+1},\ldots,y^{**}_{N})' \quad ,$$

and

$$y^{**}_{t} = y_t^{(\lambda_y)} / \sqrt{f(Z_t)} - \sum_{\ell=1}^{r} \rho_\ell \, y_{t-\ell}^{(\lambda_y)} / \sqrt{f(Z_{t-\ell})} \quad .$$

In a second step, concentration over σ^2 is achieved by writing $\partial\bar{\mathcal{L}}/\partial\sigma^2 = 0$, solving for σ^2 and replacing β by $\bar{\beta}$. The solution is:

$$\bar{\sigma}^2 = [1/(N-r)] \sum_{t=r+1}^{N} [y^{**}_t - \sum_{k=1}^{K} (\bar{\beta}_k \, X^{**}_{kt})]^2 \quad . \tag{18}$$

Substitution of $\bar{\beta}$ and $\bar{\sigma}^2$ in $\mathcal{L}$ yields the concentrated log-likelihood

$$\bar{\mathcal{L}} = -[(N-r)/2]\{\ln(2\pi) + 1\} - 1/2 \sum_{t=1+r}^{N} \ln f(Z_t)$$

$$- [(N-r)/2] \ln \bar{\sigma}^2 + (\lambda_y - 1) \sum_{t=1+r}^{N} \ln y_t \quad . \qquad (19)$$

In order to apply the iterative DFP gradient method, the derivatives of $\bar{\mathcal{L}}$ with respect to the elements of $\bar{\pi}$ have to be computed at each iteration. The analytical expressions of these derivatives are given in Liem [1979]. An initial solution must also be supplied to start the iterative process. Setting the λ_x's, the λ_z's and λ_y equal to one and the ρ's as well as the δ's equal to zero appears to be a good starting point, since the model reduces to a classical linear multiple regression.

After a first estimate of β has been found by using equation (17), a first estimate of the u's can be obtained from equation (1) and a Box-Jenkins (1970) type analysis applied to help determine the autoregressive (AR) structure of the error terms when the sample is sufficiently large. Although the actual AR structure of the estimated residuals could change between iterations, in practice the AR structure identified at the first iteration is retained for all subsequent iterations. Once the maximum has been attained, the AR structure of the estimated w's is also examined. We have made several numerical experiments and, in almost all cases, the estimated w's obtained at the final iteration appeared to behave as a pure white noise. If the sample is small, the Box-Jenkins techniques may not be very useful for identifying the AR pattern which must consequently be determined *a priori* by the experimenter.

To estimate the covariance matrix of $\hat{\pi}$, the maximum likelihood estimate of all the elements of π, the formula suggested by Berndt et al. (1974) is used. This requires computing the derivative of each $\mathcal{L}_t$ defined by equation (10) with respect to each element of π evaluated at $\hat{\pi}$. The gradient so obtained can be written as:

$$\frac{\partial \mathcal{L}}{\partial \pi} = \sum_{t=1+r}^{N} \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right)_{\pi=\hat{\pi}} \tag{20}$$

where

$$\frac{\partial \mathcal{L}_t}{\partial \pi} = (\ln y_t) \frac{\partial (\lambda_y - 1)}{\partial \pi} - \frac{1}{2} \frac{\partial \ln f(Z_t)}{\partial \pi} - \frac{w_t}{\sigma^2} \frac{\partial w_t}{\partial \pi}$$

and w_t denotes the contents of the square bracket in (10).

The covariance matrix is expressed as:

$$\text{cov} = \left[\sum_{t=1+r}^{N} \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right) \left(\frac{\partial \mathcal{L}_t}{\partial \pi} \right)' \right]^{-1}_{\pi=\hat{\pi}} . \tag{21}$$

6. APPLICATION TO A TRANSIT DEMAND FUNCTION

Let us apply our approach and procedures to the following aggregate demand function for public transit:

$$Q_t^d = D\ (P_t,\ T_t,\ C_t,\ Y_t,\ A_t) \tag{22}$$

where Q is the monthly demand for transit trips by school children in Montreal, P is the vector of prices of trips by the various modes and of

non-transportation goods, T and C are vectors of time, service levels and comfort characteristics of relevant modes, Y denotes income and A is a vector of rates of activities whose performance requires passenger transportation, such as employment, shopping, etc. The specification of the model will not be developed here. Suffice it to say that the above function belongs to the "direct demand" class in the transportation literature and is used in its linear form by the Montreal Transit Authority to explain and forecast monthly transit ridership with more than two dozen independent variables. More details on the model are given in Gaudry (1980b). The sample consists of 181 observations corresponding to the period December 1956 to December 1971.

In this application, BCT were applied to the dependent variable and to the 12 strictly positive explanatory variables in (22); the remaining 15 X variables were left untransformed. In some of the experiments, λ_y and all λ_x were assumed equal; in other experiments, three λ's were adopted: λ_y was allowed to differ from the two distinct $\lambda_{x(i)}$, $(i = 1,2)$, applied to subgroups of strictly positive variables. The first subgroup contained two activity variables — the same ones used to correct for heteroscedasticity — and the second subgroup contained the remaining 10 variables. The two Z variables, chosen among the X variables, were used separately or together in the heteroscedasticity structure defined by (5). The autoregressive structure defined by (9) contained four ρ's, namely ρ_1, ρ_3, ρ_4 and ρ_{12}. We shall comment on the main results summarized in Table I.

Reasonableness of the Stochastic Approach

As pointed out in Section 3, our maximum likelihood solution is valid only if the probability that y_t be equal to its higher or lower limit is negligeable. As further pointed out in Section 4, for positive values of λ_y, the fact that y_t may have an upper limit is of no practical consequence since the probability associated with the mass point at L tends to zero for sufficiently large values of L. Given that, for every case analyzed in Table I, the estimates of λ_y are positive and greater than two times their standard errors, our concern should therefore focus on the analysis of the probability that y_t be equal to zero.

We have computed, for each observation, the probability that y_t equal zero and we have recorded the highest probability on the last line of Table 1, for each of the ten cases considered. For most models, this highest probability is smaller than 10^{-6}; it therefore confirms the reasonableness of our stochastic approach. For case 5 and case 7, however, the highest probability is not absolutely negligeable, since it lies between 10^{-2} and 10^{-3}. Hence, in these two cases, the appropriateness of applying our maximum likelihood solution may be somewhat more doubtful.

Behavior of the Likelihood Function

As anticipated, the likelihood function increases as constraints are relaxed and the model becomes more general. For example, the sequence [1], [2], [3], [6], [8] and [9] represents models of increasing generality and the corresponding values of the log-likelihood functions increase systematically from 237.11 in [1] to 297.83 in [9]. In many cases, different initial solutions were tried out but they always led to the same maximum. Among all the values obtained for the coefficients associated with the BCT, with the corrections for serial correlation and with heteroscedasticity, only the estimation of λ_{z2} and δ_2 led to surprising results. On the one hand, cases [8], [9] and [10] yielded values of λ_{z2} of approximately 5: it is difficult to give an intuitive interpretation to such a high exponential value. One must note, however, that the value of δ_2 is correlatively rather low, thus attenuating the effect of the high exponent. On the other hand, the value of λ_{z2} shifts to -10.41 in case [7], while δ_2 jumps to 467.38. Although the likelihood function seems to behave well and repeated attempts to obtain another maximum from different initial solutions always yielded essentially the same estimates, this result is unexpected and might deserve further investigation.

Impact of Nonspherical Errors on the Functional Form

It is interesting to examine how the value of the BCT parameter varies between models [2] to [8]. In all of these models, it is assumed that $\lambda_y = \lambda_{x(1)} = \lambda_{x(2)} = \lambda$. Note that in case [2], when serial correlation and heteroscedasticity are not taken into account, λ appears to be significantly different from both zero or one, since the Student-t values are large, but the

model is almost linear in cases [6] or [7], when corrections are made for serial correlation and heteroscedasticity.

Corrections for Autocorrelation and Heteroscedasticity

If we remember that the statistic -2Λ is asymptotically distributed as chi-squared with ν degrees (where Λ is the log of the ratio of the constrained likelihood function over the unconstrained likelihood and ν is the number of constraints) and the critical chi-squared values are 3.84, 4.99, 7.82 and 9.49 for one to four degrees of freedom at the 95% confidence level, Table I suggests further considerations.

Comparing [1] to [2], the increase in the log-likelihood function reveals that it is useful to introduce a BCT. Then a comparison of [2] and [3] suggests that it is important to take account of serial correlation. An analysis of w's for this and all other cases where corrections were made for serial correlation showed that the residuals obtained at the maximum behaved as white noises except in case [6] where traces of seventh order serial correlation remained. Having taken serial correlation into account in case [3], the results of case [6] indicate that it is also important to correct for heteroscedasticity. Finally a comparison of cases [6] and [8] indicates that a more general correction of heteroscedasticity may be preferable. The comparison of cases [4] and [5] suggests, however, that, if the correction for heteroscedasticity is too rigid, the likelihood function may in fact decrease notably: hence, it is important to be careful when attempting to correct for heteroscedasticity, so as to use an appropriate correction. The choice of the variable to be used in the correction for heteroscedasticity is not indifferent either, as is suggested by the comparison of cases [6] and [7].

The Number of Functional Form Parameters

When the constraint $\lambda_y = \lambda_{x(1)} = \lambda_{x(2)}$ is relaxed in case [9], it appears that only λ_y is significantly different from both zero or one, while $\lambda_{x(1)}$ and $\lambda_{x(2)}$ are close to one. In other words, it appears that, in the example chosen, applying the BCT to the y's is more crucial than applying it to the X's; this is confirmed by looking at the results of case [10], where $\lambda_{x(1)}$ and $\lambda_{x(2)}$ have been set a priori to 1 and λ_y lies at 0.68, between 0 and 1. Note that applying the appropriate χ^2 tests to the log-likelihoods of cases [8], [9] and [10] reveals that for both [8] and [10], which are constrained versions of [9], the decrease in the log-likelihood function is not statistically significant, even at the 90% level. Hence, the hypothesis that the constraint imposed in either case [8] or [10] is verified cannot be reasonably rejected. The model corresponding to case [10] is particularly appealing since it simplifies the calculations and the interpretation of the results. Although one can definitely not conclude from this sole experiment that applying the BCT transformation to the y's may have an important impact when fitting multiple regression models but that applying the BCT to the explanatory variables is much less useful, this is surely a viewpoint that would deserve to be explored further.

Impact of Nonspherical Errors on Elasticities of Transit Demand

In addition to parameters pertaining to the functional form of the model and to the distribution of residuals, Table I also contains estimates of transit demand elasticities for a representative group of explanatory variables.

The full results, which are available from the authors, are not listed for reasons of brevity. Two measures of elasticity were calculated: the η-E measure denotes elasticities of the expected value of transit demand and is obtained as described in Section 4; the η-S measure, defined as $(\partial y/\partial X_i) \cdot (X_i/y)$ denotes elasticities calculated in the traditional fashion. Both measures use the sample mean values for the explanatory and lagged dependent variables and the η-S measure also uses the sample mean of transit demand. In this respect, there were no differences greater than .005 between η-E and η-S measures of elasticity of demand: for this reason, only the η-E measure is reported.

Moreover, taking into account the t-test statistics[8] associated with the corresponding β parameters, the elasticities are not extremely sensitive with respect to changes in BCT specification and to the introduction of serial correlation or of heteroscedasticity. The robustness of elasticities in this application is somewhat reassuring since it suggests that estimates obtained from less complete models are not so much less accurate than those obtained from more general models that they can lead the Transit Authority to completely erroneous interpretations or decisions.

CONCLUSION

We have developed an algorithm for solving linear regression models with BCT on the independent variables as well as on the dependent variable, while *simultaneously* taking into account corrections for serial correlation of several orders and for heteroscedasticity. The suggested correction for heteroscedasticity was effected with a general functional form which contains as special cases most correction schemes encountered in practice.

---

[8]. Note that, in Table 1, the t-tests reported along with the elasticities actually refer to the corresponding parameters and not to the elasticities themselves.

We had anticipated that, since the use of BCT on the dependent variable may modify error variance and even introduce heteroscedasticity, it would be important to add an appropriate compensating correction. Our application to a transit demand function confirmed this anticipation and underlined the importance of allowing for a fairly unrestrictive form of compensation. Our results also suggested that applying a BCT to the dependent variable may be more important than applying BCT to the independent variables. Finally, we found that, despite their strong influence on the estimates of the functional form of our demand function, even very general corrections for nonsphericalness of the distribution of residuals had a relatively minor impact on demand elasticity estimates of interest for transit policy impact analysis.

TABLE 1 : TRANSIT DEMAND BY MONTREAL SCHOOL CHILDREN (December 1956 - December 1971)
(part 1)

Parameters \ Constraints	[1] $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}=1$; no auto.; no hetero. (1); no hetero. (2)	[2] $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no auto.; no hetero. (1); no hetero. (2)	[3] $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no hetero. (1); no hetero. (2)	[4] $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no auto.; no hetero. (2)	[5] $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no auto; eq. pur. hetero (1); no hetero. (2)
Functional Form					
λ_y	1.00 ---	0.51 (2.67)	0.72 (3.27)	0.48 (2.18)	1.05 (4.66)
$\lambda_{x(1)}$		(-2.56)*	(-1.25)*	(-2.34)*	(0.24)*
$\lambda_{x(2)}$					
Autocorrelation					
ρ_1	0.00 ---	0.00 ---	0.14 (1.78)	0.00 ---	0.00 ---
ρ_3			-0.34 (-3.60)		
ρ_4			-0.09 (-1.06)		
ρ_{12}			0.38 (4.52)		
Heteroscedasticity					
λ_{z1}	1.00** ---	1.00** ---	1.00** ---	2.14 (1.74)	0.00 ---
λ_{z2}				1.00** ---	1.00** ---
δ_1 · TEMP (λ_{z1})	0.00 ---	0.00 ---	0.00 ---	0.01 (0.36)	2.00 ---
δ_2 · SNOWF (λ_{z2})				0.00 ---	0.00 ---
Selected Elasticities of Demand					
Real ticket fare PSCD ($\lambda_{x(1)}$)	η-E -0.38 (-2.31)	η-E -0.29 (-1.04)	η-E -0.42 (-1.75)	η-E -0.39 (-1.69)	η-E -0.15 (-0.97)
Real weekly earnings RAWEN ($\lambda_{x(1)}$)	0.91 (1.49)	0.21 (0.27)	0.27 (0.43)	0.58 (0.68)	0.71 (0.88)
Transit wait-time TW ($\lambda_{x(1)}$)	-0.07 (-0.33)	-0.39 (-0.85)	-0.28 (-0.74)	-0.19 (-0.49)	-0.16 (-0.61)
School registration RSCHP	0.44 (11.84)	0.47 (4.42)	0.37 (3.99)	0.48 (4.08)	0.41 (3.90)
Population of children TEMP ($\lambda_{z(2)}$)	0.54 (2.18)	0.60 (1.73)	0.60 (1.86)	0.39 (1.14)	0.34 (0.87)
Real retail sales SNOWF ($\lambda_{x(2)}$)	0.14 (3.74)	0.13 (2.60)	0.11 (2.34)	0.14 (2.46)	0.15 (2.56)
Log likelihood $\mathcal{L}$	237.11	241.20	268.00	253.73	223.36
Pseudo-(E)-R^2 ***	.907	.909	.934	.904	.885
Pseudo-(L)-R^2	.907	.911	.935	.923	.890
Max [Pr(y_t=0)] ****	1.14×10^{-7}	0.00	7.11×10^{-15}	0.00	9.75×10^{-3}

* Numbers in parentheses are asymptotic t-tests and an asterisk refers to the hypothesis that the parameter is equal to 1 instead of 0.
** Strictly speaking, when $\delta_{(i)} = 0$, $\lambda_{z(i)}$ is undefined (i = 1,2).
*** See part 2 of this table.
**** See part 2 of this table.

TABLE 1 (contd)
(part 2)

Parameters \ Constraints	(6) $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no hetero. (2)		(7) $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$; no hetero. (1)		(8) $\lambda_y=\lambda_{x(1)}=\lambda_{x(2)}$		(9) no constraints		(10) $\lambda_{x(1)}=\lambda_{x(2)}=\lambda$	
Functional form										
λ_y	1.00	(4.80) (0.02)*	1.22	(6.25) (1.11)*	0.80	(4.21) (-1.09)*	0.59	(2.90) (-2.06)*	0.68	(3.57) (-1.69)*
$\lambda_{x(1)}$							1.52	(1.46) (0.50)*	1.00	---
$\lambda_{x(2)}$							1.10	(1.89) (0.17)*	1.00	---
Autocorrelation										
ρ_1	0.25	(2.89)	0.19	(2.45)	0.57	(7.94)	0.50	(6.21)	0.60	(8.70)
ρ_3	-0.19	(-1.74)	-0.11	(-1.23)	-0.36	(-4.04)	-0.30	(-3.98)	-0.37	(-4.15)
ρ_4	-0.06	(-0.57)	-0.15	(-1.77)	0.13	(1.47)	-0.09	(1.02)	0.14	(1.60)
ρ_{12}	0.47	(6.36)	0.46	(6.53)	0.38	(5.28)	0.36	(4.55)	0.38	(5.32)
Heteroscedasticity										
λ_{z1}	1.07	(1.04) (0.07)*	1.00	---	1.43	(1.91) (0.57)*	1.81	(1.90) (0.85)*	1.64	(1.94) (0.76)*
λ_{z2}	1.00**	---	-10.41	(-0.53) (-9.35)*	4.71	(8.85) (6.97)*	4.69	(8.45) (6.65)*	4.88	(8.97) (7.13)*
δ_1 · TEEN (λ_{z1})	0.12	(0.48)	0.00	---	0.08	(0.64)	0.03	(0.49)	0.05	(0.56)
δ_2 · SHOWF (λ_{z2})	0.00	---	467.38	(2.73)	-0.05	(-1.56)	-0.05	(-1.52)	-0.04	(-1.51)
Selected Elasticities of Demand		η-E		η-E		η-E		η-E		η-E
Real ticket fare	PSCD ($\lambda_{x(1)}$)	-0.41	(-1.88)	-0.36	(-1.81)	-0.53	(-2.08)	-0.60	(-1.44)	-0.54	(-2.76)
Real weekly earnings	RAMEN ($\lambda_{x(1)}$)	0.82	(0.97)	0.32	(0.55)	2.18	(4.72)	1.99	(0.47)	2.32	(2.97)
Transit unit-time	TU ($\lambda_{x(1)}$)	-0.33	(-0.76)	0.14	(0.74)	-0.51	(-1.73)	-0.47	(-0.65)	-0.40	(-1.39)
School registration	ESCHP	0.34	(4.05)	0.36	(4.69)	0.30	(4.85)	0.34	(4.86)	0.32	(5.02)
Population of children	TEEN ($\lambda_{x(2)}$)	0.57	(1.38)	0.36	(1.94)	0.96	(3.16)	0.93	(0.76)	0.90	(3.74)
Real retail sales	SHOWF ($\lambda_{x(2)}$)	0.09	(1.57)	0.05	(1.01)	0.19	(3.04)	0.18	(1.69)	0.18	(2.31)
Log likelihood $\mathcal{L}$	277.28		288.98		296.09		297.83		297.45	
Pseudo-(E)-R^2 ***	.932		.927		.924		.925		.925	
Pseudo-(L)-R^2	.942		.950		.954		.955		.954	
Max [Pr(y_t=0)] ****	6.28×10^{-6}		3.22×10^{-3}		5.85×10^{-6}		7.11×10^{-15}		1.59×10^{-9}	

* See part 1 of this table.

** See part 1 of this table.

*** The pseudo-(E)-R^2 is computed using the following formula: $1 - \sum_{t=1+r}^{N} (y_t - E(y_t))^2 / \sum_{t=1+r}^{N} (y_t - \bar{y})^2$. This measure does not necessarily lie between 0 and 1. The pseudo-(L)-R^2 is computed using the following formula: $1 - \exp(\frac{2}{N}(\mathcal{L}_0 - \mathcal{L}))$. In this expression, $\mathcal{L}_0$ is the value of the log-likelihood under the assumption that the model is linear with spherical residual errors and that all β's but the constant are equal to 0. This measure necessarily lies between 0 and 1 and increases monotonically with $\mathcal{L}$. In the classical linear regression model, both the pseudo-(E)-R^2 and the pseudo-(L)-R^2 coincide with the traditional R^2.

**** The probability that y_t = 0 is estimated for each observation and the maximum value is recorded here.

REFERENCES

Amemiya, T., 1973, Regression Analysis when the Dependent Variable is Truncated Normal, Econometrica 41, 4, 997-1016.

Amemiya, T. and J.L. Powell, 1980, A Comparison of the Box-Cox Maximum Likelihood Estimator and the Nonlinear Two Stage Least Squares Estimator, technical report no. 322, Stanford University.

Appelbaum, E., 1979, On the Choice of Functional Forms, International Economic Review 20, 2, 449-458.

Beach, C.M. and J.G. MacKinnon, 1978, A Maximum Likelihood Procedure for Regression with Autocorrelated Errors, Econometrica 46, 1, January, 51-58.

Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation and Inference in Nonlinear Structural Models, Annals of Economic and Social Measurement 3, 653-665.

Berndt, E.K. and M.S. Khaled, 1979, Parametric Productivity Measurement and Choice Among Flexible Functional Form, Journal of Political Economy 87, 6, 1220-1245.

Bickell, P.J. and K.A. Doksum, 1981, An Analysis of Transformations Revisited, Journal of the American Statistical Association 76, no. 374, 296-311.

Box, G.E.P. and D.R. Cox, 1964, An Analysis of Transformations, Journal of the Royal Statistical Society B, 211-243.

Box, G.E.P. and G.M. Jenkins, 1970, Time Series Analysis Forecasting and Control (Holden-Day, San Francisco).

Buse, A., 1979, A Note on Generalized Heteroskedasticity, working paper, Economics Department, University of Alberta.

Carroll, R.J. and D. Ruppert, 1981, On Prediction and the Power Transformation Family, Biometrika 68, no. 3, 609-615.

Dagenais, M.G., 1982, The Tobit Model with Serial Correlation, Economics Letters, forthcoming.

Dagum, C. and E.B. Dagum, 1974, Construction de modèles et analyse économétrique, Cahiers de l'Institut des Sciences Mathématiques et Economiques Appliquées 8, no. 11-12.

Draper, N.R. and D.R. Cox, 1969, On Distributions and Their Transformations to Normality, Journal of the Royal Statistical Society B 31, 472-476.

Egy, D. and K. Lahiri, 1979, On Maximum Likelihood Estimation of Functional Form and Heteroskedasticity, Economics Letters 2, 2, 155-160.

Fletcher, R. and M.J.D. Powell, 1963, A Rapidly Convergent Descent Method for Minimization, Computer Journal 6, 163-168.

Gaudry, M., 1980a, Dogit and Logit Models of Travel Mode Choice in Montreal, Canadian Journal of Economics 13, 2, 268-279.

Gaudry, M., 1980b, A Study of Aggregate Bi-Modal Urban Travel Supply, Demand and Network Behavior Using Simultaneous Equations with Autoregressive Residuals, Transportation Research B 14, 1/2, 29-58.

Gaudry, M.J.I. and M.G. Dagenais, 1978, The Use of Box-Cox Transformations in Regression Models with Heteroskedastic Autoregressive Residuals, publication # 106, Centre de recherche sur les transports, Université de Montréal.

Gaudry, M.J.I. and M.G. Dagenais, 1979, Heteroscedasticity and the Use of Box-Cox Transformations, Economics Letters 2, 3, 225-229.

Gaudry, M. and M.J. Wills, 1978, Estimating the Functional Form of Travel Demand Models, Transportation Research 12, 257-289.

Glejser, H., 1969, A New Test for Heteroskedasticity, Journal of the American Statistical Association 64, 316-323.

Heckman, J. and S. Polachek, 1974, Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship, Journal of the American Statistical Association 69, 350-354.

Kau, J.B. and C.F. Sirmans, 1976, The Functional Form of the Gravity Model: a New Technique with Empirical Results, working paper, department of Finance, University of Illinois. Presented at the North American Meeting of the Regional Science Association in Toronto.

Kmenta, J., 1971, Elements of Econometrics (McMillan, New York), p. 461.

Liem, T.C., 1979, DSDM:L-1, A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals, Publication # 134R, Centre de recherche sur les transports, Université de Montréal.

Olsen, R.J., 1978, The Method of Box and Cox, A Pitfall, Working paper # 804, Institution for Social and Policy Studies, Yale University.

Park, R.E., 1966, Estimation with Heteroskedastic Error Terms, Econometrica 34, 888.

Poirier, D.J., 1978, The Use of the Box-Cox Transformation in Limited Dependent Variable Models, Journal of the American Statistical Association 73, 284-287.

Robinson, P.M., 1982, On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables, Econometrica 50, 27-41.

Rosett, R.N. and F.D. Nelson, 1975, Estimation of the Two-Limit Probit Regression Model, Econometrica 43, no. 1, 141-146.

Savin, N.E. and K.J. White, 1978, Estimation and Testing for Functional Form and Autocorrelation: A Simultaneous Approach, Journal of Econometrics 8, 1-12.

Schlesselmann, J., 1971, Power Families: A Note on the Box and Cox Transformation, Journal of the Royal Statistical Society B 33, 307-311.

Spitzer, J.J., 1976, The Demand for Money, the Liquidity Trap, and Functional Forms, International Economic Review 17, 220-227.

Spitzer, J.J., 1977, A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms, Journal of Econometrics 5, 117-128.

Theil, H., 1971, Principles of Econometrics (John Wiley & Sons Inc., New York).

Tobin, J., 1958, Estimation of Relationships for Limited Dependent Variables, Econometrica 25, 24-36.

Wang, G.H.K., W. Maling and E. McCarthy, 1981, Functional Forms and Aggregate U.S. Domestic Air Cargo Demand: 1950-1977, Transportation Research A 15, 249-256.

Welland, J.D., 1978, Cognitive Abilities, Schooling and Earnings: the Question of Functional Form, Review of Economics and Statistics 60, 622-627.

Zarembka, P., 1968, Functional Form in the Demand for Money, Journal of the American Statistical Association 63, 502-511.

Zarembka, P. 1974, Transformation of Variables in Econometrics, in P. Zarembka, ed., Frontiers in Econometrics (Academic Press, New York), ch. 3.

Zellner, A., 1971, An Introduction to Bayesian Inference in Econometrics, (John Wiley & Sons, Inc., New York).